ISBN: 978-628-95207-8-1. ISSN: 2414-6390. Digital Object Identifier: 10.18687/LACCEI2024.1.1.1571

Predictive analysis model to define behavioral patterns of landslide for early warning based on machine learning and data from hydrological and meteorological sensors in Chosica

Alvarado Jimenez, Carlos Adrian[1], Diaz Amaya, Edgar David[2], and Lo Coronado, Lyang Jazmin[3]
[1,3] Universidad Peruana de Ciencias Aplicadas, Peru, u201824455@upc.edu.pe, u201910479@upc.edu.pe
[2] Universidad Peruana de Ciencias Aplicadas, Peru, pcsiedia@upc.edu.pe

Abstract – Activations of streams, known as Landslide, are natural events that cause considerable damage to property and infrastructure, causing losses of around 5 billion dollars, which negatively impacts the economic stability of the country and the people. In this work, a predictive analysis model based on machine learning is proposed to predict the occurrence of Landslide in Chosica, Peru. The model was trained with data from hydrological and meteorological sensors and was able to identify behavioral patterns of the landslide with an accuracy of 85%. This study demonstrates that the proposed model is a viable tool that can perform an acceptable prediction rate with low error control.

Keywords-- Landslide, Machine learning, Prediction, Chosica, Model.

I. Introduction

According to the National Emergency Operations Center (COEN), during the 2021-2022 rainy season, 109 landslide events were reported nationwide, affecting more than 1.8 million people and claiming the lives of at least 47 people [1]. This problem has generated a negative impact on the stability of the country, affecting a total of 4 million people from 2017 to the current date and causing economic losses estimated at around 5 billion dollars [2]. These losses encompass the destruction of homes, roads, bridges and other infrastructure, as well as the loss of crops and the interruption of economic activity in the affected areas.

To prevent the scenario of landslide, currently, the SENHAMI (National Service of Meteorology and Hydrology of Peru) has sensors located in the highest areas of the mountains. These sensors emit signals when streams are activated due to heavy rains, which often prevents authorities from arriving in time to evacuate the population. Furthermore, these devices do not consider all the relevant variables in a natural phenomenon.

Prediction of landslide becomes complex due to the variety of variables and factors, such as topography, climate and soil characteristics, in addition to the limited accuracy and reliability of data from hydrological and meteorological sensors. Traditional approaches, such as comparing rainfall levels to river levels, may be ineffective due to lack of consideration of these variables. However, in other parts of the world, technological advances have been developed based on machine learning concepts that have managed to reduce the risks associated with weather phenomena, allowing authorities to prevent these natural disasters more effectively.

Digital Object Identifier: (only for full papers, inserted by LACCEI).
ISSN, ISBN: (to be inserted by LACCEI).
DO NOT REMOVE

Therefore, the implementation of a predictive analysis model based on machine learning trained with data collected by hydrological and meteorological sensors is proposed to identify behavioral patterns of landslide. This model would be integrated with existing systems in Peru and it is expected that with the use of appropriate algorithms and accurate data, its prediction capacity and accuracy will increase, thus addressing the limitations of current solutions.

The present work focuses on the ability to predict the occurrence of landslide through the analysis of historical records, with the aim of reducing the negative impact of these phenomena, focusing particularly on Chosica, an area that has been especially affected by these events.

II. Context

A. Supervised Learning

Supervised learning is a machine learning technique that requires a set of labeled data to train a model. The model learns from previously labeled examples and is then used to predict the label of new data [3].

B. Random Forest

Random Forest is a machine learning method used for classification, regression, and other predictive analytics tasks. It is a set of decision trees that are trained on different random subsets of the original data set. Each tree in the forest makes a decision and then a vote is taken to determine the final prediction [4].

C. Meteorological and hydrological sensors

Hydrological and meteorological sensors are devices that are used to measure different variables related to climate and water. Hydrological sensors measure variables such as water level, flow rate, water temperature and water quality. On the other hand, meteorological sensors measure variables such as air temperature, relative humidity, wind speed and direction, solar radiation and precipitation [5].

D. Area under the curve (AUC)

In terms of Machine Learning, it refers to the measure commonly used to evaluate the quality of a classification model in machine learning and the curve it refers to is the ROC curve (receiver operating characteristic curve), which is used to represent the true positive rate versus the false positive rate for different decision thresholds in classification. The AUC score is a number between 0 and 1 where 0 represents an incorrect prediction and 1 represents a correct prediction. A model with a higher AUC is generally considered better than one with a lower AUC [6].

E. Overall Accuracy (ACC)

It is a metric commonly used in machine learning to evaluate the performance of a classification model. This metric indicates the proportion of correct predictions made by the model out of the total predictions made. The ACC measures the model's ability to correctly classify samples [6].

III. Main Contribution

To address the problem raised, existing models previously implemented in similar situations were used, specifically in the field of landslide. These models were adapted to take advantage of the algorithms and technologies used, so that they could be applied to data from relevant hydrological and meteorological sensors. Among the data considered are Total Precipitation, Maximum Precipitation, Average Temperature, Minimum Temperature, Maximum Temperature, Relative Humidity, Wind Direction and Speed, Flow Rate and River Level.

The model was trained with data extracted from SENAMHI through Web Scrapping. Upon receiving current data, the model compares it with historical information stored in the database. If matches are found, a prediction of a possible landslide is generated and the prediction and warning alerts are displayed in the mobile application. Otherwise, the data is saved to the database for use in training the model.

The training data period spanned from 2017 to 2022, with a total of 52,560 records. Furthermore, for early warning prediction, training 1 hour before it occurred was considered.

A mobile application was developed that provides access to past predictions, and where users can obtain in real time the probability of occurrence of a landslide along with the associated meteorological data. The mobile application is shown in Fig. 1. The data is obtained through sensors installed by Senamhi in Peru. The application uses the Random Forest algorithm, recognized for its high precision in classifying data patterns, allowing you to accurately identify the areas that are most likely to suffer a landslide [4].

Fig. 1 Screenshot of the Variables History module of the application. Note. The statuses provided by the application are Normal and Huaico.

The following formulas are used by the Random Forest algorithm to obtain the expected value and variance. The last one is used to know how the model behaves when faced with data different from the training data. A model that presents a large amount of data generally tends to increase its variance, so it is important to find the balance of this metric.

$$E\left[\frac{1}{m}\sum_{i=0}^{m} y_i\right] = \frac{1}{m}\sum_{i=0}^{m} E(y_i) = \frac{1}{m} mu = u \quad (1)$$

$$\mathrm{var}\left[\frac{1}{m}\sum_{i=0}^{m} y_i\right] = \frac{1}{m^2}\sum_{i=0}^{m} \mathrm{var}\,(y_i) = \frac{1}{m^2} ym\sigma^2 = \sigma^2 \quad (2)$$

Where:

m = Number of decision trees

u = Mean

σ = Variance

E = Expected value

On the other hand, the AUC formula (area under the curve) helps to know the ability of a model to distinguish between positive and negative classes.

$$\int_{(0,1)} TPR(FPR)^{-1} dFPR \quad (3)$$

Where:

TPR = True Positive Rate

FPR = False Positive Rate

dFPR = Indicates an infinitesimal difference in the FPR value.

The model training was carried out in Azure machine Learning Studio, "AUTOML". The process consists of receiving the data to perform the ETL (data cleaning) process, then it is stored in the Azure data set and divided into two or more different parts. The objective is to use one part to train Machine Learning and the other part to evaluate the performance of the model. After adjusting the model parameters, it is trained, and its performance is evaluated. Subsequently, it is executed using a script developed in Python, which is programmed before training the model. Fig. 2 illustrates this structure. This automation not only significantly reduces the time and effort required to tune and validate the model, but also ensures greater accuracy and reliability in landslide predictions. The integration of AUTOML enables a more robust and scalable approach that adapts the model to various meteorological and geographical conditions, which can be crucial for effective application in different landslide-prone regions.

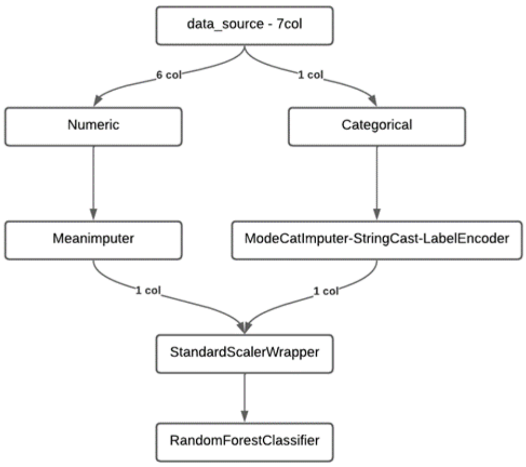

Fig. 2 Building the model in Azure ML Designer
Note The figure describes how the model is built in Azure, AUTOML.

I V. RELATED WORK

In reference to the problem described, other authors have carried out similar research with the following results.

The study by Millán C., on rainfall thresholds for landslide in the Peruvian highlands has been helpful for our research. This analysis stands out for its focus on how region-specific topographic and climatic conditions affect landslide occurrence, using advanced methods to collect and analyze data. The findings of Millán C. They not only provide scientific knowledge but also practical solutions for risk management. Our study is based on this previous research, extending it through the application of cutting-edge technology to improve the prediction and response to landslide, seeking to increase the effectiveness of early warning systems [7].

Landslide in geologically sensitive areas could present significant challenges for monitoring and early warning. The work of Yang C. addresses this problem with a technical perspective. Despite progress in landslide detection using SVM, XGBoost, and LSTM, these methods face limitations due to the variability of elements such as precipitation and changes in soil moisture content. Yang and his team propose a novel approach using deep graph learning, leveraging GNSS data to capture the temporal-spatial dynamics of landslide. Its implementation on the Wenzhou Belt Highway demonstrated that it could be more effective than previous techniques, marking an important advance in the prediction and monitoring of these phenomena [8].

The work of Yang Y. introduces a methodology based on federated learning for geotechnical risk modeling, standing out for its focus on the security and privacy of geospatial data. This proposal is of particular interest due to its ability to optimize the accuracy of landslide prediction, while facilitating collaboration between different entities without compromising the confidentiality of the information. The adoption of federated learning can be key in improving early warning systems, ensuring effective and secure data management [9].

The analysis by Marengo, on floods and landslide in Recife in 2022 sheds light on the causes behind the high mortality despite early warnings. The research highlights the confluence of inadequate disaster preparedness, poor early warning systems, and the exacerbating effects of urbanization and deforestation on the region's vulnerability. This study suggests that a significant improvement in preparedness and warning systems, together with greater collaboration between civil defense, communities and local governments, is relevant to mitigate the impact of future disasters. The methodology employed by the authors, which integrates observational data and qualitative analysis of digital media and social networks, provides a framework to understand and improve the response to disaster alerts [10].

The study by Youssef, on the Hasher-Fayfa basin employs machine learning techniques to create a susceptibility map to multiple natural hazards. The detailed methodology, ranging from data collection to modeling and validation of the risks associated with each type of natural hazard, provides a basis for understanding how these techniques can be applied in similar contexts. In our thesis, the adaptation of this approach could provide a way to more accurately evaluate and predict landslide events in Chosica, thus facilitating more informed decision-making for disaster prevention and mitigation in prone areas [11].

The research of Merghadi provides an evaluation of the performance of various machine learning algorithms in landslide prediction, using geospatial and meteorological data. This comparative analysis, which includes metrics such as AUC and accuracy, as well as cross-validation techniques to minimize overfitting, highlights the effectiveness of the Random Forest model, along with other decision tree-based models such as Extreme Gradient Boosting and Light Gradient Boosting Machine. This methodical approach could be applicable to the risk assessment of landslide in Chosica, suggesting that the incorporation of these models could improve the predictive capacity of our study. The exclusion of the KAPPA index, despite its usefulness in other research, highlights the importance of selecting performance metrics that specifically align with the goals of our predictive analysis [6].

The study by Cheng, Yu, on the application of Random Forests in landslide prediction in the Tsengwen River Basin, Taiwan, highlights the effectiveness of this method in generating spatial and precise quantitative. By identifying the main triggering factors of landslide and combining remote sensing data with geographic information systems (GIS), this approach provides relevant insights for the formulation of mitigation policies. The results, including an AUC of 98% and an overall accuracy of 91.4%, demonstrate the high reliability of the Random Forest model in this context. These findings suggest that the implementation of similar techniques could be useful to assess landslide risks in specific areas, thus supporting preparation and response to these natural events [12].

The work of Wang compares the performance of logistic regression (LR) and random forest (RF) in landslide susceptibility assessment in Yunyang, China, using geographic, topographic and geological data. The methodology included elevation data, land cover, hydrographic network and average annual precipitation, evaluated through the ROC Curve and other evaluation metrics, with a k-fold cross-validation process and stratified sampling to ensure the representativeness of the samples. The results showed a superiority of the RF model over LR, with an AUC of 0.988 versus 0.716, indicating a better ability of

RF to identify areas of high and very low susceptibility to landslide. This approach and findings could be relevant to our research, when considering the use of hydrometeorological data and the application of metrics such as AUC and ACC for model evaluation in the context of landslide prediction [13].

The paper by Singh presents a streamflow prediction model for a mountain basin in the Himalayas, using data from CMIP6 global climate models. This approach incorporated both climate and river flow data, evaluating different machine and deep learning models using six statistical measures. The Random Forest (RF) model, applied to the R3 rain scenario, showed outstanding performance, with a coefficient of determination (R2) of 0.90 and 0.78 in training and testing, respectively. This model also recorded low mean absolute error (MAE), low percentage bias (PBIAS), and high efficiency according to Kling-Gupta (KGE) and Nash-Sutcliffe (NSE) metrics, underscoring its potential for applications in resource management. water resources and the prediction of natural risks [14].

In the research of Park S., they compare and analyze the performance of 3 algorithms, random forest, decision tree, for landslide susceptibility modeling on Woomyeon Mountain in South Korea. The main result of the study was expressed in metrics and numbers, including the sensitivity, specificity, accuracy, precision, accuracy, and kappa coefficient of the three landslide models, as well as the area under the ROC curve (AUROC) for each model. The results showed that the random forest model performed the best in terms of sensitivity, accuracy and kappa coefficient in both calibration and validation data sets, while the forest rotation model performed the best in terms of specificity and accuracy in both data sets. In the present work, ROC metrics were used to evaluate the performance of the proposed model, resulting in true positive rate and false positive rate [15].

V. EXPERIMENT

In this section, the analysis and results of the experiments carried out to develop and validate the predictive analysis model for early warning of landslide will be detailed. The experiment was based on using the data collected from hydrological and meteorological sensors, to be trained with machine learning techniques and subsequently validate with a test data set, if the model is capable of identifying a landslide based on real data, to for this reason, the following experiment protocol was prepared.

A. Experimental Protocol

We present an experimental study on the occurrence of landslide, using analysis tools and several data sets. All experiments were carried out in Machine Learning Studio, a Standard_DS1_v2 virtual machine has been used, which has 1 Core 3.5 GB (RAM), 7 GB (Disk) with a cost of $0.09/hr.

For model evaluation, we use a sample of 744 data records corresponding to the month of March 2023 as a test set. These data were acquired through a Web Scraping process from the SENAMHI website, following a procedure similar to that used to obtain the training data. Subsequently, this data went through a cleaning and preprocessing process, before being fed to the predictive analysis model implemented in Visual Studio Code. The results obtained provided valuable predictions related to landslide events. According to reports from the COEN, INDECI and SENHAMI, events of landslide and torrential rains were recorded that triggered the activation of streams from March 10 to 19, in the districts of Chaclacayo, Cieneguilla, Lurigancho-Chosica and San Juan de Lurigancho, in the province of Lima.

To carry out model validation, Python was used. In this process, a specific function was created to load the necessary information, including fields such as "Date", "Time", "Flow", "River Level", "Temperature", "Precipitation", "Humidity", " Wind Direction" and "Wind Speed", using the provided data set. This feature made it easier to prepare and standardize data before entering it into the model for evaluation.

Once data preparation was complete, a web service was implemented using Python. An endpoint was created on the Azure platform, allowing remote and secure access to the model's functionality from any location with an internet connection.

With this experiment, we seek to identify the behavioral patterns of the landslide, with the purpose of predicting said event in advance.

B. Results

Of the dates on which there was a landslide, the model correctly guessed 9/9 occurrences of a landslide. However, it marked landslide on the 13th, 17th, 20th, 21st and 22nd when there were none. Additionally, the model got 26/26 correct as a “NORMAL” condition, on the days from March 1 to 10, 13, 17 to 31. Table 1 shows a fragment of the data entered with the prediction result.

TABLE I
DATA SET SAMPLE

	Variables								
Date	**Hour**	**Flow**	**River Level**	**Temperature**	**Precipitation**	**Humidity**	**Address Of the wind**	**Veil. Of the wind**	**Result**
03/12/2023	17:00:00	57.46	1	26.2	0	63	225	2.6	1
03/12/2023	00:00:00	111.10	1.3	22.9	0	0	0	0	1
03/14/2023	06:00:00	95.00	1.21	22.4	0	98	186	0	1
03/15/2023	01:00:00	115.66	1.42	22.1	0	100	265	0	1
03/17/2023	03:00:00	0.62	1.3	22.3	0	82	327	0.1	1
03/18/2023	06:00:00	56.92	1.13	twenty-one	0	95	130	0	0

After completing the training of the model, a validation process was carried out to evaluate its performance. This process consisted of providing the model with data sets that contained both past landslide events and normal events, in order to evaluate its ability to distinguish between these two classes. Additionally, more recent data that the model had not seen during its training was included, allowing for a more complete evaluation of its predictive ability in realistic situations.

Model performance evaluation was performed using a confusion matrix, which is a fundamental tool for classification analysis. This matrix breaks down the number of correct and incorrect predictions made by the model. In particular, the confusion matrix shows the number of true positives (instances of landslide correctly classified), false positives (instances that were incorrectly classified as landslide), true negatives (normal instances correctly classified), and false negatives (instances that were incorrectly classified as normal).

Fig. 3 shows this confusion matrix and the results obtained in more detail.

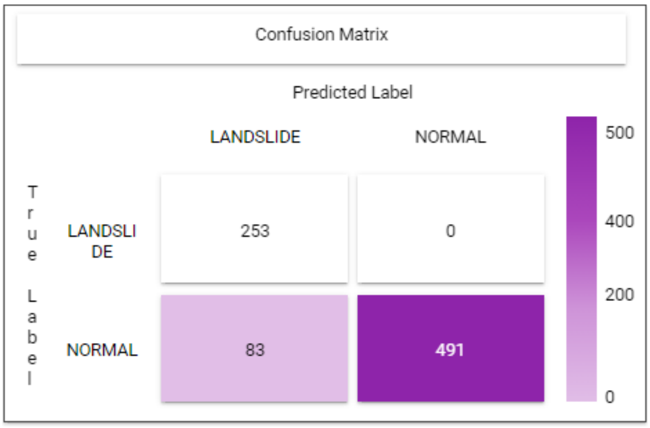

Fig. 3 Sample Data Set
Note the Figure represents the real data and those predicted by the model

The model got it right:

- 253 cases of "Landslide".
- 491 cases of "Normal".

The model confused in:

- 83 cases of "Normal" as "Landslide".

C. Metrics

The approach included validation using a confusion matrix, which allowed us to evaluate the performance of the model and its accuracy. These results suggest that the model could outperform the validation approach used by de [14]. (Wang et al., 2020) , which used k-fold cross validation and obtained a low index in its logistic regression (LR) model compared to the model proposed in this paper.

The present proposed model achieved a performance of 0.95 in the AUC metric and an accuracy of 85%. The latter surpassing the model proposed by Cheng [13]. By achieving an accuracy of 85%, the proposed model not only demonstrates its ability to make accurate predictions, but also shows an improvement of 8% compared to the previous model.

This improvement in performance is largely attributed to the quality and timeliness of the data used in training the model. Obtaining data through web scraping, one hour in advance of the landslide event, provided a significant advantage to the model by allowing a more precise and early evaluation of the conditions that lead to the occurrence of a landslide.

The detailed analysis of the model metrics, presented in fig. 4, reveals its robustness and effectiveness in predicting landslide events.

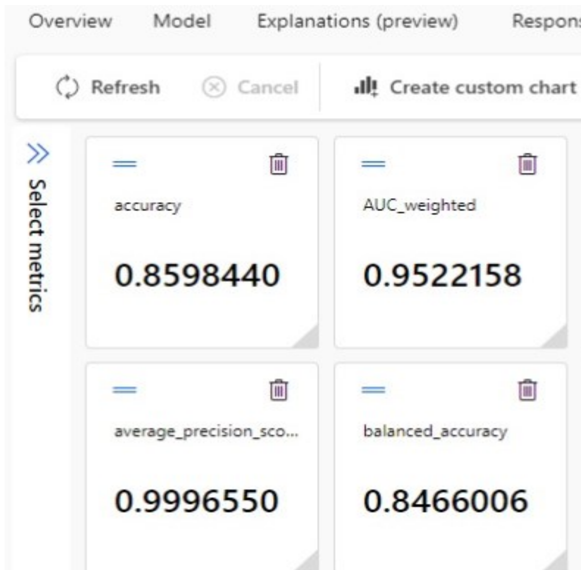

Fig. 4 Principales métricas
Nota La Figura representan las Métricas del Modelo

VI. Conclusions

The findings of this study demonstrate that the predictive analysis model based on Machine Learning is capable of predicting and contributing to early warning. The model was trained with data from hydrological and meteorological sensors, and was able to identify behavioral patterns of the landslide with an accuracy of 85 %. This accuracy was higher than previous approaches studied in the Related Work section , suggesting that the model has a greater ability to distinguish between true positives and false positives. Furthermore, validation using a confusion matrix supported the performance of the model.

The development faced two main challenges: Obtaining real-time data and historical data. The first barrier was that IoT devices were required to obtain data in real time and thus be able to provide more accurate predictions. Therefore, the Web Scrapping technique was implemented to collect data in real time from the Senamhi platform, with this, it has been possible to feed and consult the previously trained model, reaching an accuracy of 85% in the predictions. Likewise, for the second barrier, the same process of using Web Scrappping was carried out to obtain historical data from the Senamhi platform, this is because entities such as Senamhi did not provide the requested information. This process proved to be more laborious due to the time invested and the need to compile information on the specific variables required. However, once the information was collected, the model was trained using data 1 hour in advance. The data collection process using Web Scraping presents significant challenges, mainly related to variability in the quality and updating of data obtained from the SENAMHI platform. To address these limitations, rigorous data validation and cleaning protocols were implemented, ensuring that only the most accurate and up-to-date information is used for model training. Despite these efforts, reliance on external data can introduce certain uncertainties that impact model accuracy. Recognizing this limitation, it is planned to expand data collection to more regions and diversify meteorological and hydrological data sources in future iterations of the project. This could not only improve the robustness and generalization of the model, but could also facilitate the adaptation of the system to different geographical contexts and climatic conditions, which could increase its usefulness as a national landslide prediction tool.

Given the inherent complexity of landslide events and their geographic variability, it is crucial to continue expanding the scope of the study. To this end, additional research is planned to be carried out in other streams and regions, with the aim of identifying new variables that may influence the occurrence of landslide. This expansion of the data set will allow the model to be trained with a more diverse and representative sample, which will improve its ability to recognize complex patterns and make more accurate predictions in different locations.

REFERENCES

[1] COEN, & INDECI. (2023). Distribution: To the three levels of Government (National, Regional and Local). NATIONAL EMERGENCY OPERATIONS CENTER, unpublished [COEN, & INDECI. (2023). Distribución: A los tres niveles de Gobierno (Nacional, Regional y Local). CENTRO DE OPERACIONES DE EMERGENCIA NACIONAL, unpublished.]

[2] INEI. (2017). Retrieved from https://www.inei.gob.pe/media/MenuRecursivo/publicaciones_digitales/Est/Lib1467/index.html, unpublished [INEI. (2017). Recuperado de https://www.inei.gob.pe/media/MenuRecursivo/publicaciones_digitales/Est/Lib1467/index.html, unpublished]

[3] Chegini, M., Bernard, J., Berger, P., Sourin, A., Andrews, K., & Schreck, T. (2019). Interactive labeling of a multivariate dataset for supervised machine learning using linked visualizations, clustering, and active learning. Visual Informatics, 3(1), 9–17. https://doi.org/10.1016/j.visinf.2019.03.002

[4] Winham, SJ, Freimuth, RR, & Biernacka, JM (2013). A weighted random forests approach to improve predictive performance. Statistical Analysis and Data Mining: The ASA Data Science Journal, 6(6), 496–505. https://doi.org/10.1002/SAM.11196

[5] Nguyen, LD, Nguyen, HT, Dang, PDN, Duong, TQ, & Nguyen, LK (2021). Design of an automatic hydro-meteorological observation network for a real-time flood warning system: a case study of Vu Gia-Thu Bon river basin, Vietnam. Journal of Hydroinformatics, 23(2), 324–339. https://doi.org/10.2166/hydro.2021.124

[6] Merghadi, A., Yunus, A.P., Dou, J., Whiteley, J., ThaiPham, B., Bui, D.T., Avtar, R., & Abderrahmane, B. (2020). Machine learning methods for landslide susceptibility studies: A comparative overview of algorithm performance. Earth-Science Reviews, 207, 103225. https://doi.org/10.1016/J.EARSCIREV.2020.103225

[7] Millán-Arancibia, C., & Lavado-Casimiro, W. (2023). Rainfall thresholds estimation for shallow landslide in Peru from gridded daily data. Hazards Earth Syst. Sci, 23, 1191–1206. https://doi.org/10.5194/nhess-23-1191-2023

[8] Yang, C., Yin, Y., Zhang, J., Ding, P., & Liu, J. (2024). A graph deep learning method for landslide displacement prediction based on global navigation satellite system positioning. Geoscience Frontiers, 15(1), 101690. https://doi.org/10.1016/j.gsf.2023.101690

[9] Yang, Y., Lu, Y., & Mei, G. (2023). A federated learning based approach for predicting landslide displacement considering data security. Future Generation Computer Systems, 149, 184–199. https://doi.org/10.1016/j.future.2023.07.021

[10] Marengo, JA, Alcantara, E., Cunha, AP, Seluchi, M., Nobre, CA, Dolif, G., Goncalves, D., Assis Dias, M., Cuartas, LA, Bender, F., Ramos, AM, Mantovani, JR, Alvalá, RC, & Moraes, OL (2023). Flash floods and landslide in the city of Recife, Northeast Brazil after heavy rain on May 25–28, 2022: Causes, impacts, and disaster preparedness. Weather and Climate Extremes, 39, 100545. https://doi.org/10.1016/j.wace.2022.100545

[11] Youssef, AM, Mahdi, AM, Al-Katheri, MM, Pouyan, S., & Pourghasemi, HR (2023). Multi-hazards (landslide, floods, and gully erosion) modeling and mapping using machine learning algorithms. Journal of African Earth Sciences, 197, 104788. https://doi.org/10.1016/j.jafrearsci.2022.104788

[12] Cheng, Youg Sin., Yu, Teng To, Son, Nguyen Thanh. Random forests for landslide prediction in tsengwen river watershed, Central Taiwan (2021). Remote Sensing, 13(2), 199. https://doi.org/10.3390/rs13020199

[13] Wang, Y., Sun, D., Wen, H., Zhang, H., & Zhang, F. (2020). Comparison of Random Forest Model and Frequency Ratio Model for Landslide Susceptibility Mapping (LSM) in Yunyang County (Chongqing, China). International Journal of Environmental Research and Public Health, 17(12), 4206. https://doi.org/10.3390/ijerph17124206

[14] Singh, D., Vardhan, M., Sahu, R., Chatterjee, D., Chauhan, P., & Liu, S. (2023). Machine-learning- and deep-learning-based streamflow prediction in a hilly catchment for future scenarios using CMIP6 GCM data. Hydrology and Earth System Sciences, 27(5), 1047–1075. https://doi.org/10.5194/hess-27-1047-2023

[15] Park, S., Hamm, S.-Y., & Kim, J. (2019). Performance Evaluation of the GIS-Based Data-Mining Techniques Decision Tree, Random Forest, and Rotation Forest for Landslide Susceptibility Modeling. Sustainability, 11(20), 5659. https://doi.org/10.3390/su11205659